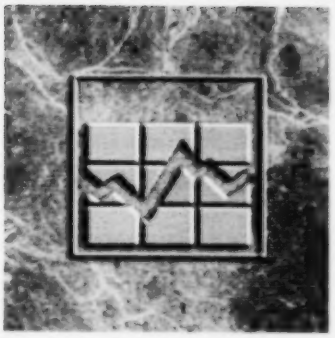
Catalogue no. 75F0002MIE — No. 003
ISSN: 1707-2840
ISBN: 978-0-662-47912-3

Research Paper

Income Research Paper Series

Sample Rotation for the Survey of Labour and Income Dynamics



by Sylvie Michaud, Lecily Hunter and Georges Lemaître

Income Statistics Canada
Jean Talon building, Ottawa, K1A 0T6

Telephone: 1-613-951-7355

Statistics Canada Statistique Canada

Canada

SAMPLE ROTATION FOR THE SURVEY OF LABOUR AND INCOME DYNAMICS

March 1992

Sylvie Michaud, Social Survey Methods Division

Lecily Hunter, Social Survey Methods Division

Georges Lemaître, Household Surveys Division

The SLID Research Paper Series is intended to document detailed studies and important decisions for the Survey of Labour and Income Dynamics. These research papers are available in English and French, at no charge. To obtain a summary description of available documents or to obtain a copy of any, please contact Philip Giles, Manager, SLID Research Paper Series, by mail at 11-D8 Jean Talon Building, Statistics Canada, Ottawa, Ontario, CANADA K1A 0T6, by telephone (613) 951-2891, or by fax (613) 951-3253.

EXECUTIVE SUMMARY

As SLID is a longitudinal survey, it is desirable to maintain respondents in the sample for as long as possible. However, sample attrition and changes in the population result in the sample becoming less representative of the population as time passes. To balance these factors, the sample for a longitudinal survey usually is comprised of panels, with each panel being representative of the target population. Starting with a fixed sample size, the paper examines feasible options for the number of panels in the sample and the length of time which each panel remains in the survey. The rationale for the selected option is reviewed.

1. INTRODUCTION

Data collection for the Survey of Labour and Income Dynamics (SLID), Statistics Canada's first longitudinal social survey, is due to begin in January 1993. The sample size has been fixed at 40,000 households. However, the sample design must achieve a compromise between many factors:

a) **longitudinal vs. cross-sectional requirements**

For longitudinal analysis, the length of time in which sampled units are kept should be as large as possible. However, the survey is expected to be able to produce some cross-sectional estimates, the reliability of which would decrease every year due to changes in the population and to sample attrition. Therefore, some sample replacement is necessary.

A "reasonably long" period is required for event history analysis as events of interest are not all going to occur conveniently in the middle of the in-sample period. A longer period improves the quality of the results obtained in modelling; and it increases the opportunities to collect additional background information. The following are among the substantive issues that cannot be effectively examined if the time frame is too short:

- completed spells of unemployment and the total amount of unemployment experienced over a span of years;
- completed episodes of poverty;
- the lead-up and aftermath of marital separation;
- determinants of worker productivity and skill mix;
- employment stability;
- impact of illness on employment and income;
- transitions in life cycle;
- wealth changes.

In the trade-off between sample size and length of time in sample, the latter is much more critical to longitudinal analysis. It is felt that five years is the minimum in-sample time.

b) **the primary uses of the data**

As mentioned, both cross-sectional and longitudinal analysis must be accommodated. In addition, some provincial and other sub-national level analysis will be required.

c) **the anticipated response rates**

The higher the nonresponse rates are expected to be, the higher the sample size is required to produce estimates of equal reliability. The impact of nonresponse on the bias of the estimates is another factor.

d) **response burden**

This is of concern not only for the impact on a given interview, but also its impact on future interviews. Sample attrition will undoubtedly occur at every time of interview as respondents decide to no longer participate.

2. WORKING ASSUMPTIONS

a) The overall sample size will be 40,000 households. However, initial SLID plans called for a 50,000 sample, and the ease or difficulty of expanding to this size at some future point is considered in the analysis.

b) The initial SLID sample will be a sub-sample selected from the Labour Force Survey (LFS). Unless there are indications that response problems result from the use of LFS samples, future panels will be selected in a similar manner.

c) SLID will include a preliminary interview and regular annual interviews. The first interview will be done as a LFS supplement. (Note that this requires the sample to be selected from LFS households, as indicated in (b) above. If a sample of non-LFS households is selected, the preliminary interview would have to be conducted separately. Its content would have to be redesigned as it will be designed to use data already collected by the LFS. The cost implications are obvious.) Thereafter, respondents would be interviewed (by telephone whenever possible) two times per year: once in January to collect labour market information (following the Labour Market Activity Survey (LMAS) model) and once in May to collect income information (using a modified version of the Survey of Consumer Finances (SCF)). (More details about the decision on interview times is available in SLID Research Paper 92-02 "Survey of Labour and Income Dynamics: Possible Interview Dates".)

d) A supplementary (to SLID) survey may be conducted each May on a sub-sample of 10,000 households.

e) The non-response and attrition rates are similar to those obtained by the U.S. Bureau of the Census for their SIPP survey (Survey of Income and Program Participation, similar in nature to SLID). The actual SIPP rates are given in Appendix A. They show a higher attrition rate for the first few panels, with an incremental attrition of 1.8% each time an interview is conducted. However, the LMAS response rate is around 92% and the SCF just under 80%. SLID may therefore experience non-response rates above those assumed here.

f) Households contain, on average, two persons eligible for the survey.

3. SAMPLE ROTATION OPTIONS

Several sample rotation plans were considered. Some were not examined at length. Under one option, the entire sample of 40,000 households would remain in sample for five years, and then completely replaced in one go. This would impose major operational requirements in years when the sample changed. As well, this design would be analytically vulnerable to the timing of changes in government programs, economic downturns and so on. Methodologically, sample deterioration would result in cross-sectional data being very much better for some years than for others. Therefore, it was concluded that some type of rotation plan involving part of the sample only was highly desirable.

No option was considered that would result in a substantially different sample size from one year to the next. For example, this would occur for a two-panel approach with a five-year in-sample period; the cross-sectional sample size would be 40,000 and 60,000 in alternate years. While longitudinally this poses no major problems, the reliability of cross-sectional data would fluctuate from year to year.

In the end, three sample rotation plans were examined in detail.

In the rotation plans presented below, P refers to the preliminary interview, while a number indicates the number of years in the sample to date (for example, 3 refers to the third year of interviewing a particular panel).

Option A: 40,000 Households with Annual Rotation of 1/5 of the Sample (8,000 Households)

Panel	1993	1994	1995	1996	1997	1998	1999	2000	2001...
A	P	1	2	3	4				
B	P	1	2	3	4	5			
C		P	1	2	3	4	5		
D			P	1	2	3	4	5	
E				P	1	2	3	4	5
F					P	1	2	3	4
G						P	1	2	3
H							P	1	2

The 1993 cross-section of the sample is implemented gradually, starting with two rotations from the LFS. The full cross-sectional sample has 40,000 households. A full longitudinal panel (for example, 1994-1998) includes 8,000 households.

Advantages:

- This plans offers 5 years of longitudinal data.
- The rotation plan is symmetrical, in that sample size and therefore interviewer workload are similar from year to year (1/6 of all interviews are in person and 5/6 are by telephone).
- For cross-sectional estimates, bias due to attrition may be partially offset by constant replenishment of the sample.

- If the sample size were increased to 50,000 households, the sampling plan could be modified easily (without major cost implications) since an LFS rotation group is about 10,500 households.

Disadvantages:

- Total non-response after five years could be around 33%, assuming attrition similar to that experienced by SIPP (stabilizing at 1.8% after the first three interviews). This would leave 10,700 persons in the longitudinal sample, which may be adequate for some national-level analyses but will not support provincial analyses.
- Even if sample attrition is not high, the limited sample in each panel would reduce the scope for detailed longitudinal analysis.
- Assuming that supplementary surveys are conducted and that these contribute to attrition, non-response at the end of five years could be as high as 40%.

Option B: <u>40,000 Households with Biennial Rotation of 20,000 Households</u>

Panel	1993	1994	1995	1996	1997	1998	1999	2000	2001...
A	P	1	2	3	4				
B			P	1	2	3	4		
C					P	1	2	3	4
D							P	1	2

This plan offers a richer longitudinal sample, but for a period of four years. Every two years, half the sample is replaced.

Advantages:

- This plan offers more data within each panel. (Note that 4 years of data would be available for 20,000 households compared with 16,000 in Option A.)
- Each respondent is contacted for a total of 5 years rather than 6, which reduces respondent burden.
- Given that a panel is 20,000 households, two LFS rotation groups could be selected, a relatively inexpensive approach.
- If there is a content change in SLID at some future date, the change can be implemented one panel at a time. Each half-sample is large enough to support some level of analysis.

Disadvantages:

- Longitudinal data would be available for a four-year period only.
- Every two years, a large number of personal interviews would be required. As such the interviewer's workload would be uneven.
- If five years of longitudinal data were essential, this rotation plan would not be symmetrical and the size of the cross-sectional sample would vary substantially from year to year.
- If the sample size were increased to 50,000 households, sample selection would probably be more costly. In fact, it seems likely that in rural areas it would be difficult to increase the sampling ratio in the Primary Sampling Units. Therefore additional clusters would have to be listed. (Another option would be to select a sub-sample of three rotation groups but this too is costly.)

Option C: <u>40,000 Households with Replacement of Half the Sample (20,000 Households) Every 2 or 3 Years, Depending on Response Rates</u>

This is a modified version of Option B. It offers the flexibility of deciding on the lifespan of a panel in 1997, based on response rates. Thus with this plan, 1997 is a decision point: if at that time response rates seem to be deteriorating too quickly to extend the life of the panel beyond four years, then sample rotation would occur every two years. However, if the response rates are stable and at an acceptable level, rotation would occur every three years, yielding six years of longitudinal data.

If response rates indicate that sample attrition is too high to warrant following the same respondents for six years, then a new sample will be introduced in 1998, and future panels will be kept in for four years. Sample B would nevertheless be followed for five years. (This is important because of the supplementary surveys.)

Sample Rotation Every Two Years

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003...
Panel											
A	P	1	2	3	4	5					
B				P	1	2	3	4			
C						P	1	2	3	4	
D								P	1	2	3
E										P	1

Sample Rotation Every Three Years

Panel	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002...
A	P	1	2	3	4	5	6			
B				P	1	2	3	4	5	6
C							P	1	2	3
D										P

Advantages:

- This plan provides a contingency, an option to determine panel length based on initial observed response rates.
- If six years are possible, interviewer workload is reduced, because a new panel (and therefore personal interviews) are required every three years instead of every two.
- For longitudinal analysis, this plan is optimal because it provides both larger panels and longer in-sample periods, relative to Options A and B.
- If content changes become necessary, they can be introduced one panel at a time, avoiding disruptions in the file layout for a single respondent. At the same time each panel is in itself large enough to support some analysis.

Disadvantages:

- If six years of annual labour and income data are collected, it means a total of seven years of interviewing which represents substantial respondent burden.

- Sample attrition may be high.
- As in Option B, if the sample were increased to 50,000 households, it would result in additional listing costs.

4. CONCLUSIONS

While SLID is required to produce certain cross-sectional data, it is first and foremost a longitudinal survey. Option C with sample rotation every three years was chosen as the most appropriate to support longitudinal analysis. As well, it allows the greatest flexibility to incorporate changes to the sample rotation if the planning assumptions turn out to be invalid.

Appendix A.

Response Rate and Sample Attrition, SIPP

(based on 1984 panel, eligible persons)

Responded to all 8 interviews		72.2%
Total Attrition		19.6%
Attrition by Wave:		
- Wave 2	4.3%	
- Wave 3	3.7%	
- Wave 4	3.0%	
- Wave 5	2.8%	
- Wave 6	2.2%	
- Wave 7	1.9%	
- Wave 8	1.8%	
Other non-response		8.1%